STA254

CORRESPONDENCE ANALYSIS AND RELATED METHODS

Michael Greenacre

Department of Statistics, Stanford University, fall 2008

Week 1: Homework exercises

Slides, supporting material and R scripts can be found at:

http://www.econ.upf.edu/~michael/stanford

Try out as much as possible of the first week's R script up to Computing Step 4 (copy-and-paste the commands but also try to understand each one: you also have available an R reference card which summarizes the main commands in R).

REMEMBER: when inputting data into R as we describe it in Computing Step 0, the instruction read.table("clipboard") must never be copied-and-pasted, because the data are in the clipboard, so would be over-written. You must physically type this instruction in the R command window.

Compare the R code in Computing Step 4 with the functions braycurt and chidist that are also given on the website, to see how to "package" these routines as functions.

An R challenge (optional – high difficulty for R beginners):

You will see that in Computing Step 4, the Bray-Curtis dissimilarity is calculated in a double 'for' loop, calculating each element of the dissimilarity matrix one at a time, whereas the chi-square distance is calculated using matrix–vector multiplications, with no sign of looping over the row and column indices. Using 'for' loops is, in fact, very inefficient in R, so we do our best to avoid them. The Bray-Curtis index can be calculated without using loops. Can you work out a way to do this?